Data Set Properties

Name: "ATP Tennis 2013 – 2023"
From: https://www.kaggle.com/datasets/dissfya/atp-tennis-2013-2023
Size: 25k+ entries
Timeframe: 2013 to April 2023
Unique Columns: 17

General summary:
The dataset is more than 25k entries of matches from the ATP tour from 2013 to 2023. The data has 17 unique columns that include the tournament name, the date a match was played, the designated series of tournament (which designates the point reward for the tournament), whether the match was played inside or outside, the round in the tournament, the scoring format of the match (best of 3 or 5), the players separated into Player_1 and Player_2 so they can be matched with their ranking or points in another two columns, the winner, the odds of each player winning the match, and the score. All of the data is contained in a csv that can be found on Kaggle under the apt name of "ATP Tennis 2013 – 2023."

Goals in place for this data:
This data can be used to find relationships between players' win/ loss ratios, their number of upsets, and other information. The ranking, point, and odds information is all numeric data while the rest of the information is qualitative data.

| Tournament | Categorical: the name of the tournament the match was played at |
|---|---|
| Date | Interval: the date when a match was played |
| Series | Catagorical: the level of the tournament. This determines how many points are received by a player for each match won. |
| Court | Catagorical: Indoor or Outdoor court |
| Round | Catagorical: the round of the tournament i.e "1st round… Semifinals" |
| Best of | Catagorical: 3 or 5 maximum sets |
| Player_1 | Catagorical: name of the first player in a match |
| Player_2 | Catagorical: name of the second player in a match |
| Winner | Catagorical: name of the winner of the match |
| Rank_1 | Numeric: the rank of player 1 |
| Rank_2 | Numeric: the rank of player 2 |
| Pts_1 | Numeric: the total points of player 1 from the calendar year |
| Pts_2 | Numeric: the total points of player 2 from the calendar year |
| Odd_1 | Numeric: the odds player 1 will win the match |
| Odd_2 | Numeric: the odds player 2 will win the match |
| score | Catagorical: the final score of the match |

Data Set 2 Properties


Name: "ATP Men's Tour"
From: https://www.kaggle.com/datasets/jordangoblet/atp-tour-20002016/
Size: 25k+ entries
Timeframe: 2000 to Sep 2014
Unique Columns: 54

General summary:
The dataset is more than 25k entries of matches from the ATP tour from 2000 to 20214. The data has 54 unique columns that include the tournament name, tournament location, the date a match was played, the designated series of tournament (which designates the point reward for the tournament), whether the match was played inside or outside, the round in the tournament, the scoring format of the match (best of 3 or 5), the players separated into winners and losers so they can be matched with their ranking, the odds of each player winning the match, and the score. All of the data is contained in a csv that can be found on Kaggle under the apt name of "ATP Men's Tour."

Goals in place for this data:
This data is much easier to use for upset data even though it is not as recent. It also has the location of each tournament so that I can put them up on a map.

Notes for Tennis Data


All data is in csv format, ready for use within standard spreadsheet applications.


Key to results data:


| ATP | Tournament number (men) |
|---|---|
| WTA | Tournament number (women) |
| Location | Venue of tournament |
| Tournament | Name of tounament (including sponsor if relevant) |
| Data | Date of match (note: prior to 2003 the date shown for all matches played in a single tournament is the start date) |
| Series | Name of ATP tennis series (Grand Slam, Masters, International or International Gold) |
| Tier | Tier (tournament ranking) of WTA tennis series. |

| | |
|---|---|
| Court | Type of court (outdoors or indoors) |
| Surface | Type of surface (clay, hard, carpet or grass) |
| Round | Round of match |
| Best of | Maximum number of sets playable in match |
| Winner | Match winner |
| Loser | Match loser |
| WRank | ATP Entry ranking of the match winner as of the start of the tournament |
| LRank | ATP Entry ranking of the match loser as of the start of the tournament |
| WPts | ATP Entry points of the match winner as of the start of the tournament |
| LPts | ATP Entry points of the match loser as of the start of the tournament |
| W1 | Number of games won in 1st set by match winner |
| L1 | Number of games won in 1st set by match loser |
| W2 | Number of games won in 2nd set by match winner |
| L2 | Number of games won in 2nd set by match loser |
| W3 | Number of games won in 3rd set by match winner |
| L3 | Number of games won in 3rd set by match loser |
| W4 | Number of games won in 4th set by match winner |
| L4 | Number of games won in 4th set by match loser |
| W5 | Number of games won in 5th set by match winner |
| L5 | Number of games won in 5th set by match loser |
| Wsets | Number of sets won by match winner |
| Lsets | Number of sets won by match loser |
| Comment | Comment on the match (Completed, won through retirement of loser, or via Walkover) |

Key to match betting odds data:

| | |
|---|---|
| B365W | Bet365 odds of match winner |
| B365L | Bet365 odds of match loser |
| B&WW | Bet&Win odds of match winner |
| B&WL | Bet&Win odds of match loser |
| CBW | Centrebet odds of match winner |
| CBL | Centrebet odds of match loser |
| EXW | Expekt odds of match winner |
| EXL | Expekt odds of match loser |
| LBW | Ladbrokes odds of match winner |
| LBL | Ladbrokes odds of match loser |
| GBW | Gamebookers odds of match winner |
| GBL | Gamebookers odds of match loser |
| IWW | Interwetten odds of match winner |
| IWL | Interwetten odds of match loser |
| PSW | Pinnacles Sports odds of match winner |
| PSL | Pinnacles Sports odds of match loser |
| SBW | Sportingbet odds of match winner |
| SBL | Sportingbet odds of match loser |
| SJW | Stan James odds of match winner |
| SJL | Stan James odds of match loser |
| UBW | Unibet odds of match winner |
| UBL | Unibet odds of match loser |

| MaxW | Maximum odds of match winner (as shown by Oddsportal.com) |
|---|---|
| MaxL | Maximum odds of match loser (as shown by Oddsportal.com) |
| AvgW | Average odds of match winner (as shown by Oddsportal.com) |
| AvgL | Average odds of match loser (as shown by Oddsportal.com) |

Tennis-Data would like to acknowledge the following sources which are currently utilised in the compilation of Tennis-Data's results and odds files.

Results:

Xscores - http://www.xscores.com/

ATPtennis.com - http://www.atptennis.com/

ATP Tour Rankings and Results Page - http://www.stevegtennis.com/

Livescore - http://www.livescore.net/

Rankings:

ATPtennis.com - http://www.atptennis.com/

ATP Tour Rankings and Results Page - http://www.stevegtennis.com/

WTA TOur Rankings - http://www.sonyericssonwtatour.com

Betting odds for matches generally represent the most recent before play starts, as reported by oddsportal.com and the individual bookmakers.